Conditional Independence and Identifiability in Graphical Models

Alexander Diaz, Luis David Garcia-Puente, Shaowei Lin,
Sonja Petrovic, Michael Stillman

Abstract

We introduce a new *Macaulay2* package `GraphicalModels`, which constructs ideals corresponding to discrete graphical models and to Gaussian graphical models used in algebraic statistics. It computes conditional independence statements in the form of trek separation and trek ideals using variants of the Bayes ball algorithm, and determines the identifiability of Gaussian graphical models.

1 Introduction

This package extends an existing package `Markov`. It is used to construct ideals corresponding to discrete graphical models, as described in several places [1].

The package also constructs ideals of Gaussian Bayesian networks and Gaussian graphical models (graphs containing both directed and bidirected edges), as described in the papers [2, 3].

Further, the package contains procedures to solve the identifiability problem for Gaussian graphical models as described in the paper [4]

2 Mathematical Background

2.1 Discrete Graphical Models

Let $X_1, \dots, X_n$ be discrete random variables where X_i takes values in the finite set $[d_i] = \{1, 2, \dots, d_i\}$. Let $p_{x_1 \cdots x_n}$ be an indeterminate representing the probability $\mathrm{Prob}(X = x), X = (X_1, \dots, X_n), x = (x_1, \dots, x_n)$, and consider the polynomial ring $\mathbb{R}[p]$ by these indeterminates.

Let A, B, and C be pairwise disjoint subsets of $\{X_1, \dots, X_n\}$ where A and B are non-empty. We say that A is independent of B given C if

$$\begin{aligned}&\mathrm{Prob}(A = a, B = b, C = c) \cdot \mathrm{Prob}(A = a', B = b', C = c)\\ &= \mathrm{Prob}(A = a, B = b', C = c) \cdot \mathrm{Prob}(A = a', B = b, C = c)\end{aligned}$$

for all values a, a', b, b', c which the random variables in A, B, C can take. Symbolically, we write this as $A \perp\!\!\!\perp B \mid C$. By marginalizing over values of random variables not in $A \cup B \cup C$, each of the terms $\text{Prob}(A = a, B = b, C = c)$ can be expanded as a sum of indeterminates $p_{i_1 i_2 \cdots i_n}$. Thus, conditional independence statements translate into homogeneous quadratic polynomials in $\mathbb{R}[p]$.

Given a directed acyclic graph G on vertices $\{1, \ldots, n\}$, let $\text{pa}(i)$ denote the set of parents of a vertex i, and $\text{nd}(i)$ the set of its nondescendents. We say that X factors according to G if

$$\text{Prob}(X = x) = \prod_{i=1}^{n} \text{Prob}(X_i = x_i \mid X_{\text{pa}(i)} = x_{\text{pa}(i)}).$$

for all values $x \in D$. It can be shown that if X factors according to G, then the following conditional independence statements are satisfied.

1. Pairwise Markov. If $i \to j$ is not an edge of G, then $X_i \perp\!\!\!\perp X_j \mid X_{\text{nd}(i)\setminus j}$.
2. Local Markov. For all vertices i of G, then $X_i \perp\!\!\!\perp X_{\text{nd}(i)} \mid X_{\text{pa}(i)}$.
3. Global Markov. If C d-separates A and B, then $A \perp\!\!\!\perp B \mid C$.

Here, we say that C d-separates (direct separates) A and B in G if every path (not necessarily directed) from A to B contains either a noncollider $i \in C$ or a collider i with $C \subseteq \text{nd}(i)$, where a vertex i is a collider if $a \to i \leftarrow b$ on the path.

Compute conditional independence statements using the Bayes ball algorithm.

2.2 Gaussian Graphical Models

Consider a mixed graph G with vertices $V = \{1, \ldots, n\}$ and edges E which could be directed $i \to j$, undirected $i - j$ or bidirected $i \leftrightarrow j$. We assume that there is a partition $U \cup W = \{1, \ldots, k\} \cup \{k+1, \ldots, n\}$ of V such that all undirected edges have their vertices in U, all bidirected edges have their vertices in W, and there are no directed edges $w \to u$ with $w \in W$ and $u \in U$.

Let Λ be an $n \times n$ matrix with $\Lambda_{ij} = 0$ if $i \to j \notin E$ and $\Lambda_{ii} = 0$ for all i. Let K and Φ be symmetric positive definite matrices, with rows and columns indexed by U and by W respectively, such that $K_{ij} = 0$ if $i - j \notin E$, $\Phi_{ij} = 0$ if $i \leftrightarrow j \notin E$ and $K_{ii}, \Phi_{ii} > 0$ for all i. Now, consider normal random variables $X_1, \ldots, X_n$ with zero means and covariance matrix

$$\Sigma = (I - \Lambda)^{-T} \begin{pmatrix} K^{-1} & 0 \\ 0 & \Phi \end{pmatrix} (I - \Lambda)^{-1}.$$

Hence, the joint distribution of these random variables is parametrized by the $\Lambda_{ij}, K_{ij}, \Phi_{ij}$ corresponding to directed, undirected and bidirected edges of G, as well as the K_{ii}, Φ_{ii} corresponding to vertices in U and W.

conditional independence coming from trek separation, they translate to rank conditions on Σ

Algorithm for computing using Theorem so and so in Seth's paper. Creating an auxiliary DAG.

2.3 Identifiability of Gaussian Graphical Models

3 Examples

Use `Graphs` package to create graphs for the models.

Here is a typical use of this package. We create the ideal in 16 variables whose zero set represents the probability distributions on four binary random variables which satisfy the conditional independence statements coming from the "diamond" graph $d \to c$, $b \to a$.

```
i1 : G = digraph  {{a,{}},{b,{a}},{c,{a}},{d,{b,c}}}
o1 = Digraph{a => set {}    }
             b => set {a}
             c => set {a}
             d => set {b, c}
o1 : Digraph
i2 : R = markovRing (2,2,2,2)
o2 = R
o2 : PolynomialRing
i3 : S = globalMarkovStmts G
o3 = {{{a}, {d}, {b, c}}, {{b}, {c}, {d}}}
o3 : List
i4 : I = markovIdeal(R,G,S);
o4 : Ideal of R
i5 : netList pack(2,I_*)
```

Sometime an ideal can be simplified by changing variables. Very often, by using, marginMap (missing documentation), such ideals can be transformed to binomial ideals. This is the case here.

```
i6 : F = marginMap(1,R)
o6 = map(R,R,{p        - p       , p        - p       , p        - p       ,
               1,1,1,1    2,1,1,1   1,1,1,2    2,1,1,2   1,1,2,1    2,1,2,1
              p        - p       , p        - p       , p        - p       ,
               1,1,2,2    2,1,2,2   1,2,1,1    2,2,1,1   1,2,1,2    2,2,1,2
              p        - p       , p        - p       , p       , p       ,
               1,2,2,1    2,2,2,1   1,2,2,2    2,2,2,2   2,1,1,1   2,1,1,2
              p       , p       , p       , p       , p       , p       })
               2,1,2,1   2,1,2,2   2,2,1,1   2,2,1,2   2,2,2,1   2,2,2,2
o6 : RingMap R <--- R
i7 : I = F I;
o7 : Ideal of R
i8 : netList pack(2,I_*)
     +-------------------------------------+-------------------------------------+
o8 = |- p       p        + p       p       |- p       p        + p       p       |
     |   1,1,1,2 2,1,1,1    1,1,1,1 2,1,1,2|   1,1,2,2 2,1,2,1    1,1,2,1 2,1,2,2|
     +-------------------------------------+-------------------------------------+
```

```
|- p        p         + p        p        |- p        p         + p        p        |
|   1,2,1,2 2,2,1,1     1,2,1,1 2,2,1,2|   1,2,2,2 2,2,2,1     1,2,2,1 2,2,2,2|
+--------------------------------------+--------------------------------------+
|- p        p         + p        p        |- p        p         + p        p        |
|   1,1,2,1 1,2,1,1     1,1,1,1 1,2,2,1|   1,1,2,2 1,2,1,2     1,1,1,2 1,2,2,2|
+--------------------------------------+--------------------------------------+
```

This ideal has 5 primary components. The first component is the one that has statistical significance. It is the defining ideal of the variety parameterized by the the factorization of the probability distributions according to the graph G. The remaining components lie on the boundary of the simplex and are still poorly understood.

```
i9 : netList primaryDecomposition I
```

The following example illustrates the caveat below.

```
i10 : H = digraph {{d,{b,a}},{c,{}},{b,{c}},{a,{c}}}
o10 = Digraph{a => set {c}   }
              b => set {c}
              c => set {}
              d => set {a, b}
o10 : Digraph
i11 : T = globalMarkovStmts H
o11 = {{{a}, {b}, {d}}, {{c}, {d}, {a, b}}}
o11 : List
i12 : J = markovIdeal(R,H,T);
o12 : Ideal of R
i13 : netList pack(2,J_*)

i14 : F = marginMap(3,R);
o14 : RingMap R <--- R
i15 : J = F J;
o15 : Ideal of R
i16 : netList pack(2,J_*)
      +--------------------------------------+--------------------------------------+
o16 = |- p        p         + p        p        |- p        p         + p        p        |
      |   1,2,1,1 2,1,1,1     1,1,1,1 2,2,1,1|   1,2,1,2 2,1,1,2     1,1,1,2 2,2,1,2|
      +--------------------------------------+--------------------------------------+
      |- p        p         + p        p        |- p        p         + p        p        |
      |   1,1,1,2 1,1,2,1     1,1,1,1 1,1,2,2|   1,2,1,2 1,2,2,1     1,2,1,1 1,2,2,2|
      +--------------------------------------+--------------------------------------+
      |- p        p         + p        p        |- p        p         + p        p        |
      |   2,1,1,2 2,1,2,1     2,1,1,1 2,1,2,2|   2,2,1,2 2,2,2,1     2,2,1,1 2,2,2,2|
      +--------------------------------------+--------------------------------------+
```

Note that the graph H is isomorphic to G, we have just relabeled the vertices. Observe that the vertices of H are stored in lexicographic order. Also note that the this graph isomorphism lifts to an isomorphism of ideals.

This package requires Graphs.m2, as a consequence it can do computations with graphs whose vertices are not necessarily labeled by integers. This could potentially create some confusion about what does $p_{i_1 i_2 \ldots i_n}$ mean. The package orders the vertices lexicographically, so $p_{i_1 i_2 \ldots i_n} = p(X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n)$ where the labels $X_1, X_2, \ldots, X_n$ have been ordered lexicographically. Therefore, the user is encouraged to label the vertices in a consistent way (all numbers, or all letters, etc). how to find the internal sorting?

-MixedGraphs
-trekSeparation
-trekIdeal
-parameter identification
-GraphicalModels.info?

Acknowledgements. Dan Grayson. Amelia Taylor.

References

[1] L. D. Garcia, M. Stillman and B. Sturmfels: Algebraic geometry of Bayesian networks, *J. Symbolic Comput.* **39** (2005) 331-355.

[2] S. Sullivant: Algebraic geometry of Gaussian Bayesian networks, *Adv. in Appl. Math.* **40** (2008) 482–513.

[3] S. Sullivant, K. Talaska, and J. Draisma: Trek separation for Gaussian graphical models, *Ann. Statist.* **38** (2010) 1665–1685.

[4] L. D. Garcia-Puente, S. Spielvogel and S. Sullivant: Identifying causal effects with computer algebra, accepted in *Proceedings of the 26th Conference of Uncertainty in Artificial Intelligence.*